Green Fluorescent Protein Tagging Drosophila Proteins at Their Native Genomic Loci With Small \( P \) Elements

Peter J. Clyne, Jennie S. Brotman, Sean T. Sweeney and Graeme Davis

Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143-0448

Manuscript received March 31, 2003
Accepted for publication August 11, 2003

ABSTRACT

We describe a technique to tag Drosophila proteins with GFP at their native genomic loci. This technique uses a new, small \( P \) transposable element (the Wee-\( P \)) that is composed primarily of the green fluorescent protein (GFP) sequence flanked by consensus splice acceptor and splice donor sequences. We demonstrate that insertion of the Wee-\( P \) can generate GFP fusions with native proteins. We further demonstrate that GFP-tagged proteins have correct subcellular localization and can be expressed at near-normal levels. We have used the Wee-\( P \) to tag genes with a wide variety of functions, including transmembrane proteins. A genetic analysis of 12 representative fusion lines demonstrates that loss-of-function phenotypes are not caused by the Wee-\( P \) insertion. This technology allows the generation of GFP-tagged reagents on a genome-wide scale with diverse potential applications.

The advent of green fluorescent protein (GFP) as a means to visualize proteins in living cells has begun a revolution in many fields of cellular biology (Chalfie et al. 1994). Three major drawbacks currently hamper experiments using GFP tagging. First, proteins of interest, once tagged, must generally be overexpressed and this overexpression can generate novel phenotypes that disrupt the processes being studied. Second, overexpression will often disrupt the fine architecture of protein localization. Finally, to tag a protein, a full-length cDNA must be available.

We have developed a GFP trap in Drosophila that is capable of generating full-length GFP fusion proteins throughout the genome. This GFP trap is based on the mobile genetic \( P \) element in Drosophila in which a GFP sequence is flanked by splice acceptor and splice donor sequences. This enables GFP to be spliced onto the native gene transcripts and take advantage of the endogenous splicing apparatus. The expression of GFP fusion proteins generated in this manner is controlled by endogenous genomic regulatory sequences. As a result, the correct spatial and temporal expression patterns are achieved and the resulting fusion proteins are not expressed at elevated levels.

Several transposable-element-based trapping methodologies have been recently pursued. Most gene traps and enhancer traps in Drosophila have been designed to facilitate gene disruption (Bier et al. 1989; Wilson et al. 1989; Lukacsovich et al. 2001). In contrast, the GFP trap that we describe here is designed to generate intact fusion proteins for studies of subcellular localization in a living cell. This GFP trap is the first described that has little apparent alteration of protein function. The protein-trapping technology that we describe here is similar to recent work from the Chia laboratory (Morin et al. 2001). A fundamental difference between these approaches resides in the construction of the mobile \( P \) element. To make the Wee-\( P \) less mutagenic than standard \( P \) elements, we have used the FLP-FRT system to remove the mini-w\(^+\) gene so that the resulting Wee-\( P \) element is small in size and lacks a second transcription unit. Our genetic data suggest that the Wee-\( P \) rarely generates loss-of-function phenotypes when homozygous or when placed in trans to deficiency chromosomes that uncover the insertion location. This allows the Wee-\( P \) to be visualized as a homozygous element, thereby tagging a larger portion of the endogenous transcript, which can be essential for visualizing proteins of low abundance. In addition, this may allow the Wee-\( P \) protein fusions to be used as reporters for changes in protein abundance during development or in response to experimental manipulation.

MATERIALS AND METHODS

Construction of the Wee-\( P \) element: The Wee-\( P \) element consists of a GFP sequence flanked by splice donor and splice acceptor sites, a mini-white gene flanked by FRT sites, an ampicillin resistance gene and origin of bacterial replication (pUC segment), and \( P \)element inverted repeat ends. To generate this transposon, we first made an intermediate vector containing the following three elements: (1) a pUC segment and \( P \)element ends cut from pCaSpeR-4, (2) a mini-white gene flanked by FRT sites cut from FLP-IMP (Keller et al. 2002), and (3) a short linker sequence. The GFP sequence was PCR amplified from pEGFP tubulin (CLONTECH, Palo Alto, CA) using the primers JB Not GFP-2 (TTAATTTACTAT...).
Molecular analysis of P-element insertions and Western blotting: The insertion site of the Wee-P element in generated lines was determined using inverse PCR, as described by the Berkeley Drosophila Genome Project (http://www.fruitfly.org/about/methods/reverse.pcr.html). PCR amplification was performed using first-round primers IPCR-4 (CAATCA TATCGGTGTCTCACG) and IPCR-6 (GATTACCCCTAGGC ATGCC) and second-round nested primers IPCR-7 (ACT ATTCCTTTTACCTCGCAG) and IPCR-9 (ACCTCTCGAGAT CGAGTCC). Western blots were done with standard SDS-PAGE gels according to standard protocols.

RESULTS

Construction and expression of the Wee-P: To generate new reagents for live visualization of proteins we developed a P-element-based strategy for tagging proteins with GFP at their native genomic loci. This approach takes advantage of the endogenous splicing apparatus utilized by the majority of genes in Drosophila: ~82% of the genes in Drosophila have introns and, of those, there are an average of 2.5 introns per gene (Long et al. 1995; Deutsch and Long 1999; M. Long, personal communication). There is also evidence that exons are frequently modular, encoding distinct motifs within a gene (Long et al. 1995). Therefore, genes with exons may tolerate the insertion of a new motif such as GFP without significant alterations to their function.

To create a small P element that would act as a GFP trap and hop frequently, we took a two-step strategy. First, we designed a P element containing the GFP sequence as well as a scorable marker (mini-white) that would allow us to identify transgenic animals harboring the P-GFP element (Figure 1A). Second, we excised the mini-white gene using the FLP recombinase system (Figure 1B; Golic and Lindquist 1989). After removing the mini-white gene from the P-GFP element, the total size of the P element was ~1.9 kb (compared to standard mini-white-containing P elements that are 10–15 kb; see Bier et al. 1989; Wilson et al. 1989). We reasoned that a smaller P element would be less mutagenic and increase the likelihood that a trapped protein would be expressed at endogenous levels by the native genomic transcriptional regulatory sequences. We refer to this P element as the “Wee-P” due to its small size. As P elements show a tendency to insert near or in the 5’ untranslated region (UTR) of genes, we chose to keep the ATG of GFP in the Wee-P so that 5’ spliced GFP fusions would be expressed (Liao et al. 2000). Therefore, we also placed consensus translation initiation sequences just 5’ to the ATG (Cavener and Ray 1991). Although sequences associated with an initial ATG were placed in the Wee-P, we deliberately chose not to include any promoter sequences so that GFP expression would accurately reflect the expression of the trapped protein. As introns occur in three phases, we constructed three different Wee-P elements appropriate for each intron phase (Figure 1C).

GGGCGCGCTTTACTAATTCAGGTTCTATATAAGCGAGA GCCTGG and JB-GFP BamHI (ATTCCTTTTACCTCGCAGATA AAGAATTTACTAATTCAGGTTCTATATAAGCGAGA GCCTGG) and inserted into this intermediate vector. The most conserved Drosophila splice acceptor site, CAG, and splice donor site, ATG, were placed in the Wee-P, we deliberately chose to trap and hop frequently, we took a two-step strategy. First, we designed a P element containing the GFP sequence as well as a scorable marker (mini-white) that would allow us to identify transgenic animals harboring the P-GFP element (Figure 1A). Second, we excised the mini-white gene using the FLP recombinase system (Figure 1B; Golic and Lindquist 1989). After removing the mini-white gene from the P-GFP element, the total size of the P element was ~1.9 kb (compared to standard mini-white-containing P elements that are 10–15 kb; see Bier et al. 1989; Wilson et al. 1989). We reasoned that a smaller P element would be less mutagenic and increase the likelihood that a trapped protein would be expressed at endogenous levels by the native genomic transcriptional regulatory sequences. We refer to this P element as the “Wee-P” due to its small size. As P elements show a tendency to insert near or in the 5’ untranslated region (UTR) of genes, we chose to keep the ATG of GFP in the Wee-P so that 5’ spliced GFP fusions would be expressed (Liao et al. 2000). Therefore, we also placed consensus translation initiation sequences just 5’ to the ATG (Cavener and Ray 1991). Although sequences associated with an initial ATG were placed in the Wee-P, we deliberately chose not to include any promoter sequences so that GFP expression would accurately reflect the expression of the trapped protein. As introns occur in three phases, we constructed three different Wee-P elements appropriate for each intron phase (Figure 1C).
Screening with the Wee-P for GFP-tagged proteins:
For our initial screen we crossed Wee-P<sup>ST0</sup> (a second chromosome phase 0 Wee-P insertion that, due to its location, does not form a fusion protein) to a stock of Δ2-3, Dr/TM6b (Figure 1D). Males heterozygous for Wee-P<sup>ST0</sup> and Δ2-3, Dr were then crossed to y w animals. Embryos were aged until late embryogenesis to allow time for zygotic protein translation before dechorionation and visualization on microscope slides. Embryos were plated as a monolayer on standard microscope slides and visualized using a fluorescent dissection microscope equipped with a ×10 compound objective. Each slide contained ~2000 embryos and was screened in 30 min. Using this method GFP+ embryos were found at a frequency of ~1/400 animals.

Each embryo identified as having GFP expression was placed in an individual vial. Adults that emerged were then crossed to y w. In some cases the Δ2-3, Dr chromosome had to be selected against. Initially, we performed IPCR directly on the progeny of the initial GFP animal. However, these animals typically had three or more independent Wee-P insertions, which complicated the completion of the IPCR step. We therefore did two rounds of outcrossing on the basis of GFP expression to clean the chromosomes. IPCR on these animals generally yielded single clear bands. We then performed sequence analysis of these single bands to determine the site and orientation of the putative fusion event. In a number of cases sequence analysis of RT-PCR products was performed to confirm the exact nature of the fusion event (Table 1).

Approximately 60% of all GFP-positive animals represent in-phase fusion events. Of the remaining 40% of animals that are selected on the basis of GFP expression, a large percentage are due to a hot spot for Wee-P insertions in the lola locus. In these animals, the GFP amino acid sequence plus a few additional nonsense amino acid residues are translated after the Wee-P exon fuses with a 5′ UTR exon of the lola locus that contains multiple stop codons. Other nonfusion events include cases where the Wee-P element has landed in the 5′ UTR of a gene whose first intron is not in the same phase as the Wee-P element.

Identity and subcellular localization of GFP fusion events: To date we have analyzed >100 lines that express GFP. We have characterized a subset of these lines in molecular and genetic detail to assess the properties of Wee-P fusion events. A diverse array of proteins have been tagged with GFP using the Wee-P element, includ-

![Figure 1](image-url)
ing transmembrane proteins as well as cytoplasmic and nuclear proteins (Figures 2–4; Table 1). The GFP fusion events can occur N-terminally as well as within the open reading frame (ORF). N-terminal fusions are derived from a variety of insertion sites including upstream of the 5′ UTR, within an exon of the 5′ UTR, and within introns bound by 5′ UTR exons. Internal fusion events are derived from insertions in introns bound by coding exons and constitute about half of the fusions (Table 1).

For internal fusion events the average size of the intron into which Wee-P elements have been observed to insert is 2.9 kb (SD = 4.1 kb). The smallest intron into which a Wee-P has inserted and generated a fusion is 183 bp. In comparison, the average intron size for fusions discovered with the protein trap transposon is ~8 kb (Morin et al. 2001), suggesting that the shorter Wee-P transposon may have a significant advantage in trapping shorter introns.

An analysis of subcellular localization for Wee-P fusions demonstrates that the fusion proteins are correctly localized in each insertion line analyzed to date. For example, the Wee-P114:GFP fusion with calmodulin is widely expressed and enriched in neuronal tissues (Figure 2A). In a second example, a fusion with Ciboulot (a G-actin-binding protein) is distributed in several different tissues, such as the central nervous system, the chordotonal organs, and the trachea (Figure 2, C–E). The subcellular localization of Wee-P3:GFP:Cib within each of these cell types is cytoplasmic, as predicted for a G-actin-binding protein. A fusion with the transmembrane protein Sec61 shows localization consistent with an endoplasmic reticulum (ER) resident protein (Figure 2F). Finally, in the Wee-P:GFP fusion with the putative chromatin remodeling protein Kismet, GFP is localized to the nucleus as predicted, where it is observed in puncta (Figure 2, G and H; Daubresse et al. 1999).

**Molecular analysis of Wee-P fusion events:** We examined a subset of Wee-P fusions by Western blots using anti-GFP antibodies. In all the fusions analyzed we found bands of the correct predicted size (Figure 3A). Furthermore, these Western blots indicate that there is generally no evidence of altered protein stability or degradation. It is also important to determine how much of the native transcript includes the GFP sequence. A previously published antibody to the Cib protein worked well on Westerns and demonstrates that a majority of the protein made from the Wee-P:3:Cib locus includes GFP (Figure 3B; Boquet et al. 2000). However, in this instance, there is a reduction in the total protein expression level. To address this question further we performed RT-PCR on five different Wee-P lines and found that in each case, in addition to transcripts that include the GFP, wild-type transcripts are also formed (Table 1).

**Table 1**

<table>
<thead>
<tr>
<th>Line</th>
<th>Gene</th>
<th>Location</th>
<th>Function</th>
<th>Predicted fusion</th>
<th>No.</th>
<th>Localization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wee-P1</td>
<td>hsc70-4</td>
<td>88E8-9</td>
<td>Chaperone</td>
<td>N terminal</td>
<td>3</td>
<td>Matches published</td>
</tr>
<tr>
<td>Wee-P2</td>
<td>kismet</td>
<td>21B3-4</td>
<td>Chromatin remodeling</td>
<td>Internal</td>
<td>2</td>
<td>Matches published</td>
</tr>
<tr>
<td>Wee-P3</td>
<td>ciboulot</td>
<td>4B1</td>
<td>Cytoskeleton</td>
<td>N terminal</td>
<td>1</td>
<td>Matches published</td>
</tr>
<tr>
<td>Wee-P4</td>
<td>sec61</td>
<td>26D4-7</td>
<td>ER channel</td>
<td>Internal</td>
<td>1</td>
<td>Matches published</td>
</tr>
<tr>
<td>Wee-P5</td>
<td>lar</td>
<td>98C3</td>
<td>Unknown</td>
<td>N terminal</td>
<td>1</td>
<td>Cytosolic</td>
</tr>
<tr>
<td>Wee-P6</td>
<td>belle</td>
<td>85A5-8</td>
<td>RNA helicase</td>
<td>1, N terminal; 1, internal</td>
<td>2</td>
<td>Matches published</td>
</tr>
<tr>
<td>Wee-P20</td>
<td>CG9894</td>
<td>23A7-B1</td>
<td>Unknown</td>
<td>Internal</td>
<td>5</td>
<td>See text</td>
</tr>
<tr>
<td>Wee-P25</td>
<td>hrb98DE</td>
<td>98D3-4</td>
<td>RNA binding</td>
<td>N-terminal truncation</td>
<td>2</td>
<td>Matches published</td>
</tr>
<tr>
<td>Wee-P26</td>
<td>mhc</td>
<td>36A8-9</td>
<td>Muscle cytoskeleton</td>
<td>Internal</td>
<td>1</td>
<td>Matches published</td>
</tr>
<tr>
<td>Wee-P27</td>
<td>tropomyosinI</td>
<td>88E9-12</td>
<td>Muscle cytoskeleton</td>
<td>Internal</td>
<td>1</td>
<td>Matches published</td>
</tr>
<tr>
<td>Wee-P28</td>
<td>qfp6</td>
<td>ND</td>
<td>Ribosomal protein</td>
<td>N terminal</td>
<td>1</td>
<td>Matches published</td>
</tr>
<tr>
<td>Wee-P59</td>
<td>palp2</td>
<td>44A7-8</td>
<td>RNA polyadenylation</td>
<td>Internal</td>
<td>1</td>
<td>Matches published</td>
</tr>
<tr>
<td>Wee-P114</td>
<td>calmodulin</td>
<td>48F8-10</td>
<td>Calcium binding</td>
<td>N terminal</td>
<td>6</td>
<td>Matches published</td>
</tr>
<tr>
<td>Wee-P118</td>
<td>cfp1</td>
<td>50D2-3</td>
<td>Cystein protease</td>
<td>Internal</td>
<td>1</td>
<td>Matches published</td>
</tr>
<tr>
<td>Wee-P123</td>
<td>GH03576</td>
<td>67F4-68A7</td>
<td>Unknown</td>
<td>N terminal</td>
<td>1</td>
<td>Cytosolic</td>
</tr>
<tr>
<td>Wee-P153</td>
<td>topoisomeraseI</td>
<td>13B6</td>
<td>DNA maintenance</td>
<td>Internal</td>
<td>1</td>
<td>Matches published</td>
</tr>
<tr>
<td>Wee-P157</td>
<td>CG7970</td>
<td>65A5</td>
<td>Peroxisomal membrane</td>
<td>N terminal</td>
<td>1</td>
<td>Cytosolic</td>
</tr>
<tr>
<td>Wee-P283</td>
<td>ATP1palhA</td>
<td>93B</td>
<td>Na pump</td>
<td>N terminal</td>
<td>2</td>
<td>Plasma membrane</td>
</tr>
<tr>
<td>Wee-P338</td>
<td>Pyk</td>
<td>94A10</td>
<td>Pyruvate kinase</td>
<td>N terminal</td>
<td>1</td>
<td>Cytosolic</td>
</tr>
<tr>
<td>Wee-P345</td>
<td>Karybeta3</td>
<td>82D</td>
<td>Nuclear protein import</td>
<td>Internal</td>
<td>1</td>
<td>Cytosolic</td>
</tr>
</tbody>
</table>

For an explanation of “Predicted fusion,” see text.

* Lines for which RT-PCR has confirmed the predicted fusion and the presence of wild-type transcript.

# Summary of selected Wee-P fusions

- **Wee-P** fusion with **kismet** (a G-actin-binding protein) is distributed in several different tissues, such as the central nervous system, the chordotonal organs, and the trachea (Figure 2, C–E).
- The subcellular localization of **Wee-P3:GFP:Cib** within each of these cell types is cytoplasmic, as predicted for a G-actin-binding protein. A fusion with the transmembrane protein **Sec61** shows localization consistent with an endoplasmic reticulum (ER) resident protein (Figure 2F).
- Finally, in the **Wee-P:GFP** fusion with the putative chromatin remodeling protein **Kismet**, GFP is localized to the nucleus as predicted, where it is observed in puncta (Figure 2, G and H; Daubresse et al. 1999).

## Molecular analysis of Wee-P fusion events

We examined a subset of Wee-P fusions by Western blots using anti-GFP antibodies. In all the fusions analyzed we found bands of the correct predicted size (Figure 3A). Furthermore, these Western blots indicate that there is generally no evidence of altered protein stability or degradation. It is also important to determine how much of the native transcript includes the GFP sequence. A previously published antibody to the Cib protein worked well on Westerns and demonstrates that a majority of the protein made from the Wee-P:3:Cib locus includes GFP (Figure 3B; Boquet et al. 2000). However, in this instance, there is a reduction in the total protein expression level. To address this question further we performed RT-PCR on five different Wee-P lines and found that in each case, in addition to transcripts that include the GFP, wild-type transcripts are also formed (Table 1).
Figure 2.—Wee-P fusions show a wide variety of expression patterns. (A) Wee-P114:GFP:Cam is found at the neuromuscular junction in third instar larvae, as well as in the motoneuron axons and in the muscle nuclei. (B) Wee-P114:GFP:Cam is found in the soma and dendrites of the third instar sensory neurons of the peripheral nervous system. (C) Wee-P3:GFP:Cib is expressed throughout the optic lobes of the larval central nervous system and in a subset of cells within the ventral ganglion. (D) Cytoplasmic Wee-P3:GFP:Cib expression is observed within the chordotonal organs of the third instar larva. (E) Cytoplasmic Wee-P3:GFP:Cib expression is observed in the tracheal system of the third instar larvae. (F) Wee-P4:GFP:Sec61 is localized in patches of the cytoplasm of cells. Depicted are epithelial cells of a third instar larva. (G) Wee-P2:GFP:Kis expression is widespread and localized to the nuclei, as has been previously characterized for this gene. Expression of Wee-P2:GFP:Kis is shown in an imaginal disc at low magnification. (H) At higher magnification, discrete puncta of Wee-P2:GFP:Kis fusion proteins in the nuclei are visible. Such puncta were noted in all cell types examined.

Figure 3.—Wee-P fusion proteins are of the predicted size and can reveal expression patterns for uncharacterized genes. (A) Representative fusion lines were blotted and probed with anti-GFP antibodies. Single clear bands are seen in all cases and are approximately the correct predicted sizes. (B) An analysis of the Wee-P3:GFP:Cib fusion in homozygous Wee-P3 flies with an anti-Cib antibody indicates that the majority of the protein made from the locus is tagged with GFP. There is a reduction in the total amount of protein made from this locus in comparison to yw controls. (C) Wee-P20 is a fusion with the uncharacterized but annotated gene CG9894. Nuclei from third instar muscle nuclei are shown.

an important role in vesicle fusion at the neuromuscular junction (NMJ; BRONK et al. 2001). RT-PCR using primers specific for GFP and the open reading frame of Hsc70-4 indicates that a fusion mRNA is made and demonstrates that the native splicing apparatus recognizes the splice donor sites in the Wee-P (Figure 4A). Additional RT-PCR using primers designed to GFP and the exon composed of the 5’ UTR of Hsc70-4 also indicates that the 5’ UTR exon is linked to this new mRNA and demonstrates that the native splicing apparatus also recognizes the splice acceptor sequences in the Wee-P (Figure 4A). The fusion product between GFP and Hsc70-4 should be a perfect fusion of GFP onto the Hsc70-4 protein without any additional alterations to Hsc70-4. Sequence analysis of our RT-PCR product indicates that such a perfect fusion has occurred. Western analysis of Wee-P1 flies using an anti-GFP antibody indicates that a protein of the predicted ~98 kD is formed (Figure 3A). A genetic analysis of Wee-P1, described in detail below, indicates that the Wee-P1 insertion does not alter Hsc70-4 function.

We have examined the tissue localization pattern of Wee-P1:GFP:Hsc70-4 in larvae and found a broad and low-level expression of GFP in all cells and a strong expression in tissues undergoing morphological changes, such as third instar imaginal discs (data not shown). Previous studies with RNA in situ have revealed that Hsc70-4 is expressed in virtually all cells, with enrichment in tissues undergoing extensive rapid growth and changes in shape (PERKINS et al. 1990). Thus the Wee-P1:Hsc70-4 fusion provides direct evidence that GFP fusions from Wee-Ps can reflect the correct tissue localization patterns.

Analysis of Wee-P1 also underscores that the correct subcellular localization can occur in Wee-P fusions. Prior studies have demonstrated that Hsc70-4 is found...
both in the cytoplasm and in the nucleus (Carbajal et al. 1993) and that such a distribution for GFP is observed in Wee-P1 (Figure 4C). Recent research has also demonstrated that Hsc70-4 is enriched at the third instar neuromuscular synapse where it is involved in vesicle cycling (Bronk et al. 2001). In Wee-P1, GFP expression is enriched at the NMJ and, within the NMJ, Wee-P1:GFP:Hsc70-4 is localized to hot spots (Figure 4B).

This fusion also demonstrates appropriate trafficking of a Wee-P:GFP fusion protein. In response to heat shock, Hsc70-4 translocates to the nucleus (Carbajal et al. 1993). We subjected Wee-P1 animals to heat shock and examined the fluorescence intensity of GFP in muscle nuclei in animals before and after heat shock. We observed a 400% increase in nuclear GFP fluorescence intensity following heat shock (Figure 4, C and D). To control for nonspecific effects on protein translation and translocation, we performed a similar analysis with Wee-P114 and Wee-P4. In neither case did heat shock alter GFP fluorescence intensity. Thus, the Wee-P1:GFP:Hsc70-4 fusion traffics correctly in response to heat shock stimulation.

Wee-P fusions can be used to probe the expression patterns of uncharacterized genes: Our data examining Wee-P fusions with previously characterized genes indicate that fusion proteins are stable and behave normally within their endogenous cellular environment. On the basis of these data, it is reasonable to suspect that fusions with novel and previously uncharacterized genes may also reflect correct subcellular localization. For example, we have isolated a fusion with Larp, a large (150-kD) protein that is largely uncharacterized. The Larp protein has a conserved La domain that is most closely related to the yeast Sro9p gene (Chauvet et al. 2000). Sro9p is primarily cytoplasmic and thought to bind mRNAs and modulate their translation (Sobel and Wolin 1999). We observe that Wee-P5:GFP:Larp is widely expressed in the embryo and larval stages and within cells is localized to the cell cytoplasm (data not shown). As a second example, Wee-P20 is a fusion with the gene CG9894, which encodes a protein of unknown function. This fusion is abundantly expressed and highly localized to the nucleus in a wide variety of cell types although not in all cell types (Figure 3D).

Wee-P insertions, in general, do not disrupt the genes to which they fuse: For the Wee-P fusion events to be most useful, it is important that the presence of the Wee-P element not disrupt the function of the gene to which the fusion has occurred. We have attempted to address these issues by undertaking a genetic analysis of a representative subset of Wee-P fusion lines (Table 2). For nine of the fusion events that we analyzed genetically, there was no observable phenotype in animals homozygous for the Wee-P insertion. In three cases we performed our phenotypic analysis on the chromosome bearing the Wee-P element placed in trans over either a deficiency that uncovered the locus or a null allele of the fused gene. In all nine cases none of the predicted phenotypes were observed. Thus the absence of phenotypes in this random sampling of Wee-P lines suggests that Wee-P insertions are largely nonmutagenic.

For some of the genes for which we have fusion events,
<table>
<thead>
<tr>
<th>Line</th>
<th>Gene</th>
<th>Published phenotype</th>
<th>Test for Wee-P phenotype</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wee-P1</td>
<td><em>hsc70-4</em></td>
<td>Visible</td>
<td>Viable over Df(3R)Δ356;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>no visible phenotype</td>
<td>BRONK et al. (2001)</td>
</tr>
<tr>
<td>Wee-P2</td>
<td><em>kismet</em></td>
<td>ND</td>
<td>Lethal</td>
<td>Daubresse et al. (1999)</td>
</tr>
<tr>
<td>Wee-P3</td>
<td><em>larp</em></td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>Wee-P4</td>
<td><em>belle</em></td>
<td>Fertility</td>
<td>Lethal</td>
<td></td>
</tr>
<tr>
<td>Wee-P26</td>
<td>mhc</td>
<td>Visible</td>
<td>Lethal</td>
<td></td>
</tr>
<tr>
<td>Wee-P27</td>
<td><em>tropomyosinI</em></td>
<td>ND</td>
<td>Lethal</td>
<td></td>
</tr>
<tr>
<td>Wee-P59</td>
<td><em>pabp2</em></td>
<td>Electrophysiology</td>
<td>Lethal</td>
<td>Spradling et al. (1999); Sigrist et al. (2000)</td>
</tr>
<tr>
<td>Wee-P114</td>
<td><em>calmodulin</em></td>
<td>Visible</td>
<td>Lethal</td>
<td>Spradling et al. (1999); Nelson et al. (1997);</td>
</tr>
<tr>
<td>Wee-P118</td>
<td><em>cp1</em></td>
<td>Fertility and visible</td>
<td>ND</td>
<td>Gray et al. (1998)</td>
</tr>
<tr>
<td>Wee-P283</td>
<td>ATPalpha</td>
<td>ts viability; bang sensitivity</td>
<td>Lethal</td>
<td>Schubiger et al. (1994); Feng et al. (1997)</td>
</tr>
<tr>
<td>Wee-P338</td>
<td>Pyk</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>Wee-P345</td>
<td>Karybeta3</td>
<td>Lethal</td>
<td>Homozygous viable; no</td>
<td>Spradling et al. (1999)</td>
</tr>
</tbody>
</table>

ND indicates that there are no published hypomorphic or null alleles for the gene.

*“No visible phenotype” indicates that homozygous Wee-P flies have no observed morphological or obvious behavioral phenotypes. “Homozygous viable” means that flies can be homozygous for the Wee-P chromosome.*

There are several possible explanations for the lack of observable phenotypes in the examples presented above. First, the added GFP exon may be inert. In the literature there are now many examples of GFP-tagged proteins that are fully functional (e.g., Sweeney and Davis 2002). Alternatively, since both wild-type and GFP incorporated transcripts are usually generated (Table 1), normal function may be contributed by the wild-type transcript. However, this may not account for the lack of phenotype in examples where weak hypomorphic phenotypes should be observed such as Hsc70-4 (see above). A list of the Wee-P lines examined genetically, as well as what is known about any alleles of the fused genes, is given in Table 2.

## DISCUSSION

We have developed a broadly applicable technique for systematically tagging proteins with GFP on a genome-wide scale. Importantly, the tagged proteins are
the endogenous transcriptional and translational controls. Fusion proteins generated with this technology show the predicted tissue-level expression pattern and the correct subcellular localization allowing microdomains of protein localization to be visualized in a living organism. We anticipate that this technology will enable a new generation of experiments to be pursued in Drosophila. It may prove possible to use these reagents to study protein localization, trafficking, turnover, concentration, and translation within an in vivo genetic system without the caveat of protein overexpression that often precludes reliable experimental interpretation.

Wee-P technology is most significant if a tagged protein retains all of the features of tissue expression and subcellular localization for the native protein, and if the fusion protein functions normally. Our analysis of Wee-P1::GFP:Hsc70-4 and other fusion lines demonstrates that Wee-Ps can accurately reflect the expression patterns of the genes they tag (Figures 2 and 4). In addition, our findings that none of the representative 12 Wee-P lines analyzed have phenotypes suggests that Wee-Ps are generally not mutagenic. Thus, Wee-P technology should allow for rapid and accurate assessments of the temporal and spatial patterns of genes.

Although ~40% of the GFP-expressing lines collected in the initial screen are not true fusions, in the great majority of cases this fact can be discovered after just the IPCR step. This is because the insertion site of the Wee-P element can immediately illustrate if a true protein-trapping event has occurred or if some other event, such as fusion with a UTR exon, has occurred. Thus, determining whether or not a GFP-expressing line is a true fusion can be rapidly determined without a time-consuming molecular analysis.

A database of GFP-tagged proteins with Wee-P technology can have many potential applications. First, GFP-tagged proteins can reveal new patterns of gene expression for novel genes, as exemplified by Wee-P20::CG9894, as well as previously characterized genes, as exemplified by the tracheal and chordonal expression of WeeP:Cib (Figures 2 and 3). Second, these reagents allow for subcellular protein localization to be characterized in detail in the living animal. Third, fusion events can tag specific splice variants of a gene, allowing for detailed analysis that is otherwise difficult. Such isoform-specific tagging has occurred in our two Wee-P::GFP::Kis lines, in which only the larger isoform has been tagged. Fourth, the Wee-P will tag proteins regardless of their size, enabling fusion events with large proteins to be generated, which are otherwise difficult reagents to create. Our Wee-P::GFP::Kis lines, which have tagged a 17-kb open reading frame, are also good examples of this point. Finally, patterns of GFP expression in specific cell types can be reagents to study developmental processes such as axon guidance, synaptic plasticity, or tracheal development (see Wee-P3::GFP::Cib and Wee-P1::GFP::Hsc70-4 in Figure 2 and Figure 4, respectively).

Several important considerations must be taken into account when using Wee-P fusions. Most important is the question of the retention of wild-type protein function in the fused lines. Our phenotypic analysis of 12 of the Wee-P lines indicated no alteration to function and suggests that Wee-P insertions are largely nonmutagenic. However, we anticipate that some fusion events will alter protein function, mostly in those cases where the GFP disrupts a critical domain of the protein. Although exons are frequently modular, encoding distinct motifs within a gene (Long et al. 1995), there are instances to the contrary.

Our goal is to generate as many GFP-tagged proteins as possible with Wee-P elements. We anticipate making these reagents freely available so that other workers can perform similar screens for genes in their area of interest, and we are constructing a website to facilitate widespread use of this technology.

We thank G. Ruiz and S. Gandhi for helpful assistance. P.I.C. was supported by a postdoctoral fellowship from the Medical Investigation of Neurodevelopmental Disorders Institute at the University of California, Davis. S.T.S. is a Wellcome Prize Traveling Fellow (058327/Z/99/Z). This work was supported by a Burroughs Wellcome Young Investigator Award, a Merck Scholar Award, and a National Institutes of Health grant (44908-32574) to G.W.D.

LITERATURE CITED


Feng, Y., L. Heynh, K. Terayasu and D. M. Fambrough, 1997 The Drosophila Na,K-ATPase a-subunit gene: gene structure, pro-
Communicating editor: K. Anderson